

CLAIMS

What is claimed is:

- 1 1. A computer-implemented method for identifying optimal allocations of computing
 2 resources in a data processing arrangement having a plurality of computing machines that
 3 host a plurality of application processes, comprising:
 4 establishing a plurality of server models, each server model including one or more
 5 server nodes, wherein each server node has an associated set of capacity attributes;
 6 designating a layered relationship between the server models, wherein for a first
 7 server-model layer immediately above a second server-model layer, the second server-
 8 model layer includes respective models that represent the nodes in the first server-model
 9 layer;
 10 establishing a plurality of service models, each service model including one or
 11 more service nodes, wherein each service node has an associated set of demand attributes;
 12 designating a layered relationship between the service models, wherein for a first
 13 service-model layer immediately above a second service-model layer, the second service-
 14 model layer includes respective models that represent the nodes in the first server-model
 15 layer; and
 16 generating an optimized mapping of service nodes in a user-selected service model
 17 to server nodes in a user-selected server model as a function of the demand and capacity
 18 attributes.
- 1 2. The method of claim 1, further comprising:
 2 monitoring, while the applications processes are executing, levels of demand for
 3 computing resources that correspond to selected ones of the demand attributes;
 4 storing the levels of demand; and
 5 generating an alternate optimized mapping of service nodes in a user-selected
 6 service model to server nodes in a user-selected server model using the stored levels of
 7 demand and the capacity attributes.

- 1 3. The method of claim 2, further comprising:
2 establishing one or more service-node relationships between selected pairs of the
3 service nodes, wherein each service-node relationship has an associated transport demand
4 attribute specifying a quantity of communication resources required for communication
5 between the associated pair of service nodes;
6 establishing one or more server-node relationships between selected pairs of the
7 server nodes, wherein each server-node relationship has an associated transport capacity
8 attribute specifying a quantity of communication resources available for communication
9 between the associated pair of server nodes; and
10 generating the optimized mapping as a function of the service-node relationships
11 and server-node relationships.
- 1 4. The method of claim 3, wherein each service node has an associated set of capacity
2 attributes and further comprising generating an optimized mapping of service nodes in a
3 first user-selected service model to service nodes in a second user-selected service model
4 as a function of the demand attributes of the first service model and capacity attributes of
5 the second service model.
- 1 5. The method of claim 4, wherein each server node has an associated set of demand
2 attributes and further comprising generating an optimized mapping of server nodes in a
3 first user-selected server model to server nodes in a second user-selected server model as a
4 function of the demand attributes of the first server model and capacity attributes of the
5 second server model.
- 1 6. The method of claim 5, further comprising:
2 representing the service models and server models in XML; and
3 generating an allocation matrix in XML that represents the optimized mapping.
- 1 7. The method of claim 1, further comprising:
2 establishing one or more service-node relationships between selected pairs of the
3 service nodes, wherein each service-node relationship has an associated transport demand

4 attribute specifying a quantity of communication resources required for communication
 5 between the associated pair of service nodes;
 6 establishing one or more server-node relationships between selected pairs of the
 7 server nodes, wherein each server-node relationship has an associated transport capacity
 8 attribute specifying a quantity of communication resources available for communication
 9 between the associated pair of server nodes; and
 10 generating the optimized mapping as a function of the service-node relationships
 11 and server-node relationships.

1 8. The method of claim 7, wherein each service node has an associated set of capacity
 2 attributes and further comprising generating an optimized mapping of service nodes in a
 3 first user-selected service model to service nodes in a second user-selected service model
 4 as a function of the demand attributes of the first service model and capacity attributes of
 5 the second service model.

1 9. The method of claim 1, wherein each service node has an associated set of capacity
 2 attributes and further comprising generating an optimized mapping of service nodes in a
 3 first user-selected service model to service nodes in a second user-selected service model
 4 as a function of the demand attributes of the first service model and capacity attributes of
 5 the second service model.

1 10. The method of claim 9, wherein each server node has an associated set of demand
 2 attributes and further comprising generating an optimized mapping of server nodes in a
 3 first user-selected server model to server nodes in a second user-selected server model as a
 4 function of the demand attributes of the first server model and capacity attributes of the
 5 second server model.

1 11. The method of claim 1, wherein each server node has an associated set of demand
 2 attributes and further comprising generating an optimized mapping of server nodes in a
 3 first user-selected server model to server nodes in a second user-selected server model as a
 4 function of the demand attributes of the first server model and capacity attributes of the
 5 second server model.

1 12. An apparatus for identifying optimal allocations of computing resources in a data
2 processing arrangement having a plurality of computing machines that host a plurality of
3 application processes, comprising:

4 means for establishing a plurality of server models, each server model including
5 one or more server nodes, wherein each server node has an associated capacity attribute;

6 means for designating a layered relationship between the server models, wherein
7 for a first server-model layer immediately above a second server-model layer, the second
8 server-model layer includes respective models that represent the nodes in the first server-
9 model layer;

10 means for establishing a plurality of service models, each service model including
11 one or more service nodes, wherein each service node has an associated demand attribute;

12 means for designating a layered relationship between the service models, wherein
13 for a first service-model layer immediately above a second service-model layer, the second
14 service-model layer includes respective models that represent the nodes in the first server-
15 model layer; and

16 means for generating an optimized mapping of service nodes in a user-selected
17 service model to server nodes in a user-selected server model as a function of the demand
18 and capacity attributes.

1 13. A system for identifying optimal allocations of computing resources in a data
2 processing arrangement having a plurality of computing machines that host a plurality of
3 application processes, comprising:

4 a model repository including a plurality of server models and a plurality of service
5 models, each server model including one or more server nodes and each server node
6 having an associated set of capacity attributes, each service model including one or more
7 service nodes and each service node having an associated set of demand attributes,
8 wherein the server models are defined in a layered relationship and for a first server-model
9 layer immediately above a second server-model layer, the second server-model layer
10 includes respective models that represent the nodes in the first server-model layer, and the
11 service models are defined in a layered relationship and for a first service-model layer

12 immediately above a second service-model layer, the second service-model layer includes
 13 respective models that represent the nodes in the first service-model layer; and
 14 an optimization engine coupled to the model repository, the optimization engine
 15 configured to generate an optimized mapping of service nodes in a user-selected service
 16 model to server nodes in a user-selected server model as a function of the associated
 17 demand and capacity attributes.

1 14. The system of claim 13, further comprising:
 2 means for monitoring, while the applications processes are executing, levels of
 3 demand for computing resources that correspond to selected ones of the demand attributes;
 4 means for storing the levels of demand; and
 5 wherein the optimization engine is further configured to generate an alternate
 6 optimized mapping of service nodes in a user-selected service model to server nodes in a
 7 user-selected server model using the stored levels of demand and the capacity attributes.

1 15. The system of claim 13, further comprising:
 2 wherein the model repository further includes one or more service-node
 3 relationships between selected pairs of the service nodes, each service-node relationship
 4 having an associated transport demand attribute that specifies a quantity of communication
 5 resources required for communication between the associated pair of service nodes;
 6 wherein the model repository further includes one or more server-node
 7 relationships between selected pairs of the server nodes, each server-node relationship
 8 having an associated transport capacity attribute that specifies a quantity of
 9 communication resources available for communication between the associated pair of
 10 server nodes; and
 11 the optimization engine is further configured to generate the optimized mapping as
 12 a function of the service-node relationships and server-node relationships.

1 16. The system of claim 13, wherein each service node has an associated set of
 2 capacity attributes and the optimization engine is further configured to generate an
 3 optimized mapping of service nodes in a first user-selected service model to service nodes

4 in a second user-selected service model as a function of the demand attributes of the first
5 service model and capacity attributes of the second service model.

1 17. The system of claim 13, wherein each server node has an associated set of demand
2 attributes and the optimization engine is further configured to generate an optimized
3 mapping of server nodes in a first user-selected server model to server nodes in a second
4 user-selected server model as a function of the demand attributes of the first server model
5 and capacity attributes of the second server model.

10010929-1